Emotion Detection using Machine Learning for Students Enrolled in Online Courses

SHORT SYNOPSIS

Submitted in partial fulfillment of the requirement of the degree of

DOCTOR OF PHILOSOPHY

to

Manav Rachna International Institute of Research and Studies
(Deemed to be University)

by

Anand Kumar
22/Ph.D/S/025

Under the supervision of

Supervisor
Dr. Rashmi Agrawal
Professor, School of Computer Application

Verified and found O.K. All expected items/ components of short synopsis is well arranged and displayed properly.

24.02.2024

SCHOOL OF COMPUTER APPLICATION
MANAV RACHNA INTERNATIONAL INSTITUTE OF RESEARCH & STUDIES
SECTOR 43, SURAJKUND – DELHI ROAD, FARIDABAD 121001

SCHOOL OF COMPUTER APPLICATION

Certificate

I hereby declare that the short synopsis Title "EMOTION DETECTION USING MACHINE LEARNING FOR STUDENTS ENROLLED IN ONLINE COURSES" being submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy in Computer Application under School of Computer Application of Manav Rachna International Institute of Research & Studies, Faridabad during the academic year 2023-24. It is a bona fide record of original work carried out under the supervision of Dr. Rashmi Agrawal and has not been submitted in full or part to any other university or institute for degree or diploma. Further, it is certified that short synopsis is plagiarism free and proof has been submitted.

Anand Kumar
22/PHD/025

Signature of the Supervisor
Dr. Rashmi Agrawal, Professor, SCA

(Chairperson, DRC)

Dean
School of Computer Application

ABSTRACT

Soon after the COVID-19 outbreak, the entire education industry faced a whole new challenge for creating a secondary type of education that could be feasible, accessible, keeping in view the imposed restrictions. Online education in its various forms was introduced, and all of a sudden, the whole industry shifted to online education. The Universities/Colleges/Institutions were all of a sudden compelled to adopt online learning. Along with this, many online courses came into existence and mainly in the form of "Recorded Video" in which the student has to go through a self-paced course. MOOCs and various online courses were introduced and the same is being continued with a vast number of students enrolling in this. The study of the effectiveness of these courses is one of the most important tasks, as its effectiveness and completeness will shape its future. The study of Students' Emotion through text has to be analyzed in order to get their view on the online course they have enrolled. As per one of the research "Only 10% of enrolled students in MOOC's courses completes the course". Thus it becomes necessary to study the reason behind it. The comments by the students below each video may be analyzed using various machine learning algorithms in order to analyze students' emotion and act accordingly. This will also facilitate the instructor to improve themselves according to the outcome supplied by students in terms of text comments. Using various machine learning techniques the comments of the students may be analyzed for the effectiveness and efficiency of respective online course. There are different machine learning methods for emotion detection, namely: Random Forest (RF), Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree (DT) etc.

All of these machine learning algorithms, as mentioned above, analyze emotions in text form, extract the most representative set of features, and then perform classification or clustering to model an optimal approach.

The research will utilize cutting-edge machine learning algorithms, including natural language processing (NLP), to capture and interpret emotional cues exhibited by students.

Key objectives include the design and implementation of a robust data collection framework, the development of feature-rich models for emotion classification, and the validation of the proposed system's efficacy across diverse online learning platforms.

Keywords: Machine Learning, Online Learning, Machine Learning in Education

CONTENTS

SNo.	Description	Page No.
1	Introduction	5-6
2	Literature Review	6-13
3	Research gap and problem statement	13-14
4	Description of broad area	14-20
5	Research objectives	20
6	Research methodology	20-21
7	Expected outcome of the research	22
8	References	22-25
9	Plagiarism Report	26-30

1. Introduction

COVID-19 brought a whole new set of challenges worldwide. This was a time to explore alternative methods due to imposed lockdowns. Among these challenges, the education industry also faced significant hurdles. Traditional schooling was no longer feasible due to the pandemic, prompting the education Industry to seek alternative mechanisms. Thus, the industry swiftly transitioned from traditional teaching methods to online education. Everyone was compelled to embrace this change, as there were no viable alternatives. It was during this time that the traditional education system shifted to an online mode. This new approach was unfamiliar to both students and teachers, but there was no other option. Teachers encountered numerous new challenges, and students also faced issues with the new educational system.

The education industry began utilizing online education, with online video lectures gaining popularity. Many online courses emerged on various platforms, including MOOCs, to meet the growing demand for online learning. These courses gained popularity due to their self-paced learning and flexibility in terms of location, timing, and cost. However, online courses, including self-paced and pre-recorded video courses, had several limitations. One major limitation was understanding the emotions of students about the course. The high dropout rate of online courses indicates a need to study why students are leaving. Machine learning offers various algorithms for educational data mining to understand student emotions. Analysing student comments (in text form) is crucial for improving the quality of learning, as it allows for modifications based on student feedback. Understanding student emotions is essential for the survival and future of online education.

Different machine learning methods for emotion detection exist, as shown in Figure 1. These algorithms analyse emotions in text form, extracting the most representative features and then performing classification or clustering to reach conclusions. Emotion detection involves extracting features from the text, constructing character representations, applying

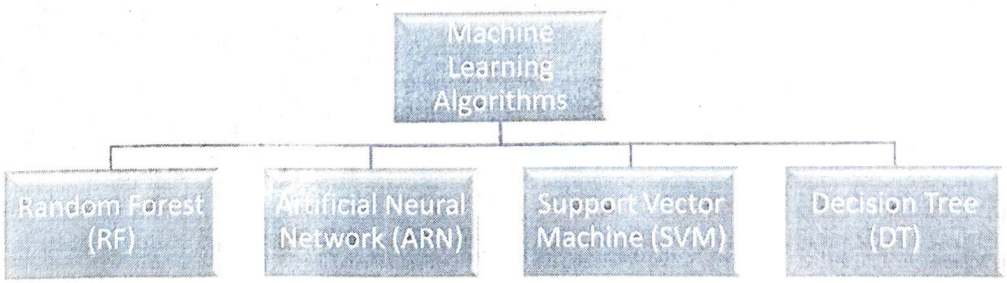

Figure 1: Machine Learning Algorithms

classifiers such as Decision Trees, Naïve Bayes, or Support Vector Machines, tokenizing the text, converting it into tokens, identifying emotional words, and using these words as keywords for classifying emotions.

Machine Learning is gaining popularity due to its significant impact on education. It allows for the analysis of student behaviour, paving the way for improvement for educators and organizations, benefiting both students and educators alike. Machine learning is effective in education for several reasons:

a) Enhancing e-learning through feedback

b) Predicting student performance

c) Providing unbiased academic evaluations of students

2. Literature Review

Various researchers have conducted studies on the topic of "Machine Learning in Education," but relatively few have focused specifically on "Emotion Detection Using Machine Learning for Online Courses." This area still offers ample scope for research, with many aspects yet to be explored. To comprehend the impact and advantages of employing machine learning in education, an extensive literature review has been undertaken, with the following papers being examined:

Wu 2020 [1] worked to find factors that influence student's performance and to find out most optimal classification method. The researcher used K-nearest neighbours, Logistic regression, Decision Tree, Boosting Tree & Random forest algorithms for his research purpose. A total 1000 rows of student performance data with 8 columns was used as dataset for carrying out the research. In the study it was found that the factors that influence student's performance in descending order are: lunch, test preparation, race, gender and parental education level.

Limitation: This study focuses only on the factors influencing student's performance in education.

Prasad et al 2022 [2] worked on implanting machine learning based google teachable machine in early childhood education. The researcher used machine learning for his research purpose. Thirty four students aged 5-6 from intermediate school, Kolkata was used as dataset for carrying out the research. In the student it was found that as a result of ML

based assessment, teachers, students, and parents are constantly informed of the student's progress toward their educational goals.

Limitation: The study focuses only on ML based teachable machines as well as no study was for Online Education.

Machova et al 2023 [3] worked on finding emotion based on text analysis. The researcher objective was to establish effective communication in machine human communication. For this purpose the text needs to be analysed and corresponding replies. The primary task which was taken into consideration was to detect human emotion and accordingly the chatbot will communicate with human in terms of text. CNN was used to create a model that will train the chatbot. The developed model was found suitable for analysing emotion in terms of text.

Limitation: The research was based on developing a model which can be integrated with chatbot. The model do not provide any scope for emotion detection using text.

Bharti et al 2022 [4] worked on detecting emotion based on text using the concept of deep learning. The research objective was to apply various DL & ML techniques to identify the algorithm that gives the most accurate result. The researcher compared various DL, ML algorithms with hybrid model (CNN+BiGRU+SVM) and it was discovered that the hybrid model performed well than the other pre-existing models for emotion detection.

Limitation: Real-time text based emotion system may be developed to facilitate the real time feedback and detection.

Liu et al 2022 [5] worked on automated detection of emotional and cognitive engagement in MOOC discussion to predict learning achievement. The primary research objective was to identify how are the different types of emotions and cognitions in the MOOC discussion identified accurately by the BERT-CNN model in interpretable manner and what are the relationships among emotional engagement, cognitive engagement and learning achievement? BERT-CNN Model technique was used to conduct the study. 8867 Learners text data on MOOC learner's platform https://www.icourse163.org (China's large online course provider) was used to conduct the study. In the research, it was found that emotional and cognitive engagement have an interactive relationship and a combined effect on learning environment.

Limitation: Student's emotion in prospects of only emotional and cognitive development has been detected. Emotion detection in terms of student's interest in the course has not been studied.

Tao 2022 [6] worked towards an understanding of the engagement and emotional behaviour of MOOC students using sentiment and semantic features. The primary research objective was to study during COVID behaviours of students in the MOOC environment in terms of sentiment/stress behaviours. The researcher used Naïve Bayes Classification, Random Forest & Deep Learning Artificial Neural Network technique for the research study. AdelaideX MicroMasters dataset was used in the research. In the research it was found that majority of students failed in the course in 2020 indicating that there were major impacts on the students in the students in their offline life.
Limitation: Emotion detection in terms of student's interest in the course has not been studied.

Edalati et al 2022 [7] worked towards the potential of machine learning algorithms for sentiment classification of students' feedback on MOOC. The primary research objective was to conduct a study to evaluate various machine learning models for aspect based opinion mining to address the challenge while shifting from offline to online mode. Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT) techniques were used in the study. 15 different computer science course on "Coursera" online learning platform were studied to carry out the research. In the research it was found that despite network optimization for the 1D-CNN and BERT model, the performances achieved in the deep learning models were less than conventional models.
Limitation: Student's emotion was detected to know the impact while shifting from offline education to online education. Emotion detection in terms of student's interest in the course has not been studied.

Bernius et al 2022 [8] worked towards machine learning based feedback on textual student answers in large courses. The primary research objective was to reduce assessment efforts on textual exercises for instructors while scaling feedback for large courses and to design a system that automatically assesses textual exercises. CoFee (Computer aided Feedback for textual exercise), a framework to generate and suggest computer aided feedback for textual exercises based on ML. 2200 records from Technical University of Munich from

2019-2023, was used to conduct the study. In the study, it was found that CoFee significantly reduces the assessment work and help the instructors deliver consistent feedback and reduce student complaints.

Limitation: The research focuses on feedback on assessment and not overall course emotion detection.

Abdi 2023 et al [9] worked towards students feedback analysis model using deep learning-based method and linguistic knowledge for intelligent educational systems. The primary research objective was to automate Student feedback analysis. DTLP (Deep Learning and teaching process) that employs Convolutional Neural Network (CNN), Bidirectional LSTM (BiLSTM) technique was used for the research. For the research purpose student feedback was collected from student surveys administered by a university in Netherlands. In the research it was found that DTLP performed the best across all methods. It got higher performance than one that learnt from a feature subset.

Limitation: Emotion detection using text for online courses has not been done.

Baashar et al 2021 [10] worked on predicting student's performance using machine learning methods: A Systematic Literature Review. The primary research objective was to find out most efficient machine learning technique for emotion detection. This was a literature review that used Artificial Neural Network (ANN), Decision Tree and Support Vector Machine technique to carry out the research. In the research online data source from Science Direct, ACM Digital Library, IEEE Explore, Scopus ad Google Scholar. As per the study, accuracy of the ML algorithms from Highest to Lowest are:

a. ANN (98.3%)
b. Decision Tree (98.2%)
c. K-nearest Neighbour (95.8%)
d. Support Vector Machine (91.3%)
e. Naïve Bayes (90%)

Limitation: In the research efficiency of the algorithms has been carried out. Emotion detection in terms of student's interest in the course has not been studied.

Alslaity et al 2022 [11] worked on emotion detection using machine learning for the current state, challenges and future directions. A total of 123 papers were reviewed in this paper to investigate research trends with machine learning. As per the paper, it was found that

interest towards emotion detection using machine learning has gained attention in recent past and SVM & Naïve Bayes are the most popular algorithms for emotion detection.
Limitation: The paper do not consider any of the paper related to online education.

Mozhdelhi et al 2023 [12] worked on identifying emotion from text for transfer learning. In the paper the performance comparison for two models i.e. BERT & RNN was done. The primary objective was to develop more accurate model to detect emotion from text. An architecture was developed using BERT model called EmotionalBERT, which will analyse emotion using text.
Limitation: In some cases the result of the model was not as expected that needs to be further analysed.

Vijayanand et al 2020 [13] worked on automatic facial expression recognition system. Facial expression recognition is widely being used in robot control, driver state recognition, computer human interaction & many more. The researcher aim to develop more robust system to ensure dynamics of facial expression. It has made efforts to study nearly all recent past papers and figures out most relevant models for FER.
Limitation: The paper was focused on facial expression and not on text based emotion.

Onyema et al 2022 [14] worked on challenges and prospects faced in academic forecasting using machine learning. The researcher worked on exploring various factors for i.e. challenges for academic forecasting. ML has the capability to understand student's weaknesses and strength along with the reason behind it. It was discovered that ML empowers teachers as well as educational institutions to know their students in a better way and to strengthen them on their weak points.
Limitation: The research was limited to explore the capabilities of ML.

Ahmad et al 2021 [15] worked on artificial intelligence and its significance/role in education. The researcher has analysed the significance of artificial intelligence in education and the need of adopting new tech and trends in education sector. The outcome of the research depicts that artificial intelligence has influenced many sector as well as education sector. AI may be used to solve many education related issues in an efficient manner.

Limitations: The research was based on importance of AI and not with analysing the text using AI & ML.

The Summary of the research papers in the literature review section has been given the Table 1 below:

Table 1 Summary of Literature Review

S.No	*Name of Author(s)*	*Method/Technique*	*Result*	*Research Gap*
01	J. Wu, 2020	K-nearest neighbours, Logistic Regression, Decision Tree, Boosting Tree, Random Forest	The factors in descending order are Lunch Test Preparation, Race, Gender and Parental Education Level.	This study focuses only on the factors influencing student's performance in education.
02	P. Yogendra Prasad, Dr. Dumpa Prasad etc.| 2022	Machine Learning	As a result of ML based assessment, teachers, students, and parents are constantly informed of the student's progress toward their educational goals.	The study focuses only on ML based Teachable machines as well as no study was for Online Education.
03	Kristína Machová etc. | 2023	Naïve Bayes (NB), Support Vector Machine (SVM)	A detection model was developed for emotion detection on neural networks.	The model lacks on detection sarcasm. Additionally, it was done keeping in view simulation with chat bot and not as per student's prospect.
04	Santosh Kumar Bharti etc.| 2022	Machine Learning & Deep Learning	The developed hybrid model [ML+DL] has better accuracy in detecting emotion.	Lack of text based emotion detection for student's emotion in online course.
05	Sannyuya Liu etc. | 2022	BERT-CNN Model	Emotional and Cognitive engagement have an interactive relationship and a combined effect on learning environment.	Student's emotion in prospects of only emotional and cognitive development has been detected. Emotion

				detection in terms of student's interest in the course has not been studied.
06	Xiaohui Tao | 2022	Naïve Bayes Classification, Random Forest, Deep Learning Artificial Neural Network	Majority of students failed in the course in 2020 indicating that there were major impacts on the students in the students in their offline life.	Emotion detection in terms of student's interest in the course has not been studied.
07	Maryam Edalati etc. | 2021	Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT)	Despite network optimization for the 1D-CNN and BERT model, the performances achieved in the se deep learning models were less that conventional models.	Student's emotion was detected to know the impact while shifting from offline education to online education. Emotion detection in terms of student's interest in the course has not been studied.
08	Jan Philip Bernius etc. | 2022	CoFee (Computer aided Feedback for textual exercise) , a framework to generate and suggest computer aided feedback for textual exercises based on ML.	CoFee significantly reduces the assessment work and help the instructors deliver consistent feedback and reduce student complaints.	The research focuses on feedback on assessment and not overall course emotion detection.
09	Asad Abdi etc. | 2023	DTLP (Deep Learning and teaching process) that employs: Convolutional Neural Network (CNN) & Bidirectional LSTM (BiLSTM)	DTLP performed the best across all methods. It got higher performance than one that learnt from a feature subset.	Emotion detection using text for online courses has not been done.
10	Yaha Baashar, gamal Alkawsi | 2021	Artificial Neural Network (ANN), Decision Tree, Support Vector Machine	Accuracy from Highest to Lowest is: ANN (98.3%), Decision Tree (98.2%), K-nearest Neighbour (95.8%), Support Vector Machine (91.3%) & Naïve Bayes (90%)	In the research efficiency of the algorithms has been carried out. Emotion detection in terms of student's interest in

				the course has not been studied.
11	Alaa Alslaity etc. | 2022	Machine Learning	After analysing 123 paper, it was found that SVM & Naïve Bayes are the most popular emotion detection models.	Textual sentiment from online courses was not taken into consideration.
12	Mahsa Hadikhah Mozhdehi etc. | 2023	BERT RNN Based Models	An architecture was developed on BERT model called EmotionalBERT for emotion detection using text for small dataset. The new model performed well in the case of transfer learning.	The developed model was developed for transfer learning and not for online learning.
13	Vijayanand G. etc. | 2020	Euclidean Distance method. Artificial Neuro-Fuzzy Inference System (ANFIS)	A detailed summary of different models and aspects of FER were presented.	The paper was focused on facial expression and not on text based emotion.
14	Edeh Michael Onyema etc. | 2022	Artificial Neural Network (ANN), Bayesian Neural Network (BNN), K-nearest neighbour (KNN)	ML empowers teachers as well as educational institutions	The research was limited to explore the capabilities of ML.
15	Sayad Fayaz Ahmad etc. | 2021	Theoretical explanations was done instead of using any algorithms	AI empowers education sector and provides significant improvement in shaping student's future.	The research was completely theoretical.

3. Research Gap & Problem Statement

After going through extensive Literature Review, the following Gaps were identified:

1. Lesser research has been done for emotion detection for online courses and there's still gap to identify the same for online course.
2. Many research model is based on retrospective analysis of emotion. Research gap exist in real time emotion detection that can provide instant feedback to the educators for the betterment of the course.

3. Many machine learning algorithms showed less accuracy rate while dealing with emotion detection using text.
4. There are different learning platforms with different features and users. It is important to focus on generalization of machine learning algorithms for different learning platforms.

Problem Statement

Current sentiment analysis models lack nuanced emotion detection in online educational content, especially within textual comments on course videos. Bridging this gap requires advanced natural language processing techniques to accurately discern and respond to diverse emotional cues. Developing a specialized emotion detection framework for online courses is essential for enhancing adaptive learning systems and fostering improved student outcomes.

4. Description of Broad Area

There are many languages in the world using which we communicate with each other. These languages are just a way to express our thought, emotion and even requirements using that particular language. Our thoughts may be of different sentiments like positive, negative or neutral. Let's take an example of an online platform selling online courses and the student may buy it and complete the course online with certification. In their website they have also provided the customer feedback section where they can express their thought regarding the course in terms of quality and services provided during the course. The user may review in a number of manners like:

User Review 1: The course was well structured and complete in all respect.

User Review 2: The course was not well structured.

User Review 3: The course content was somehow good but it would be better if you provide more practice questions for practice.

If we analyse the review as above the first review seems to be positive and the student is happy with the course structure. The second review is negative and hence the company need to look into the course structure. The third review doesn't signify whether the student is happy or unhappy with the course and hence this may be placed in "neutral" category.

There are different types of Sentiment Analysis which is extracted from set of text and that may be positive, negative and neutral. Even, this also goes beyond this and specific feelings like emotions (angry, happy, sad etc.), intentions (interested, not interested etc.) and many more. Depending upon the requirement the analysis may be done. Broadly, sentiment analysis may be divided into following four categories:

a) Aspect based sentiment analysis
b) Fine grained sentiment analysis
c) Intent based sentiment analysis
d) Emotion detection

a. Aspect based sentiment analysis

In this type of analysis the focus majorly lies on extracting opinion from the text. This is more enhanced version of traditional sentiment analysis which goes beyond only positive, negative and neutral. The statement like the paper quality of this book is not good refers to a negative aspect because the customer is not satisfied with the paper quality of the book. It may be used in different applications like:

i. Finance: Analysis of growth, risks, financial statement etc.
ii. Healthcare: Analysis of quality of care, treatment satisfaction level etc.
iii. E-commerce: Analysis of like, dislike, quality, design, price comparison etc.

b. Fine grained sentiment analysis

It uses lexicon approach for deeper sentiment on intensity, emotions, target etc. It may be used in different applications like:

i. Product review: Analyse whether product like or disliked by customer.
ii. Customer feedback analysis: Identifying customer preferences
iii. News & Media: It uses generally 5 start rating which starts using from 5 stars referring very good and 1 star referring very bad.

c. Intent based sentiment analysis

Intent based sentiment analysis is done when we have to analyse whether the text is a query, command, request, order, complaint etc. It may be used in different applications like:

i. Healthcare: To analyse patient's experience
ii. Human resource: To analyse employee feedback to know areas of improvement

d. Emotion detection

This is the emotion detection analysis using which we can go deep into the emotion detection that may be happiness, sadness, frustration, anger and many more. It is used to analyse the emotional state of the writer. It may be done using lexicons (words to convey emotion) or various machine learning algorithms. It may be used in different applications like:

i. Social Media: To know about insights of the product or service
ii. Customer feedback analysis: To identify customer behaviour to improve the product or service

Emotion detection and taking corrective decision according to the emotion for the improvement of online courses will enable the teacher as well the students to complete the course with ease. Emotion detection refers to machine learning methods that are used to analyse emotions like love, sadness, frustration, fear, anger etc. There is one more similar term i.e. sentiment analysis or opinion mining. Sentiment analysis is also a kind of text classification method which in turn returns the sentiment positive, negative or neutral after extracting subjective information from the text. Emotion detection and sentiment analysis seems to be same but emotion detection is broader prospect of sentiment analysis. In order to detect emotion from text (in our case, it will be comments on the videos) some certain set of steps are used. The steps wise process to detect emotions through text may be broadly categorized as shown in figure 2 below:

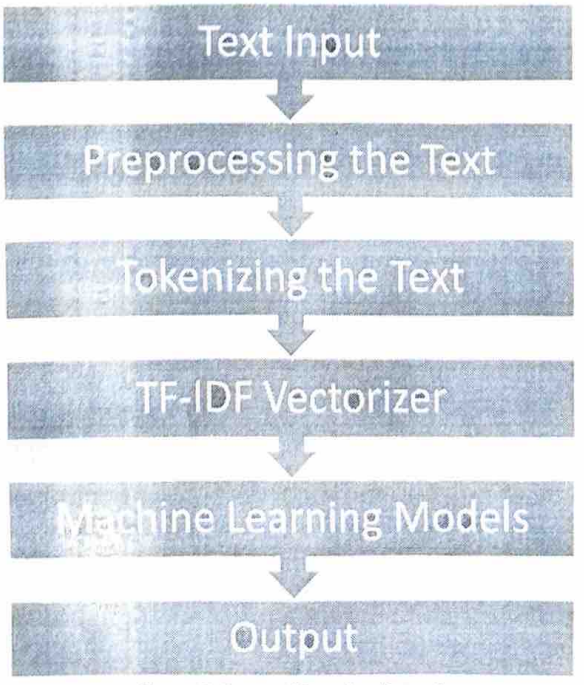

Figure 2: Steps of Emotion Detection

1. Text Input

   This is the very first step towards emotion detection. The first thing that is taken care is about the data source in which research has to be carried out. This is the raw data upon which the emotion detection will be carried out. Once the data source is there, the second step can be initiated.

2. Pre-processing the text

   In this step, refining of the text to its best form is done. Pre-processing of text includes removing punctuation marks, has tags, links, extra spaces, This also includes the process of removing stop words as well as combining similar meaning words (lemmatizing) and reduce the word to its original form (stemming the text). Once the pre-processing is complete, the next step may be taken up.

3. Tokenizing

   In this step, the next step is to break down the sentences, paragraph into words so that the same could analyse the same word by word.

4. TF-IDF Vectorizer

   In this step, use TF-IDF vectorizer to transform text to feature vectors i.e. a kind of input for the models which will analyse the sentiment. Using TF-IDF vectorizer, transform input text into features.

5. Applying Machine Learning Models

   In this step, different machine learning Models may be applied in order to detect the emotion out from the text. The following are mostly used machine learning models:

   a. Artificial Neural Network (ANN)
   b. Decision Tree (DT)
   c. K-nearest Neighbour (KNN)
   d. Support Vector Machine (SVM)
   e. Naïve Bayes

a. Artificial Neural Network (ANN)

   Artificial Neural Network (ANN) are algorithms (computational models) that are designed similar to human brain system [figure 3]. In human brain system there are enormous neurons that reacts as per the environment. In the same patter ANN's has been designed having neurons that are connected to each other using layers of network.

   ANN Architecture

   ANN consists of three layers:

   a. Input Layer: Accepts input

b. Hidden Layer: It is an intermediate layer between input & output layer.

c. Output Layer: The input is processed in hidden layer and finally the results Output.

The architecture of ANN may be represented using Figure 3 below:

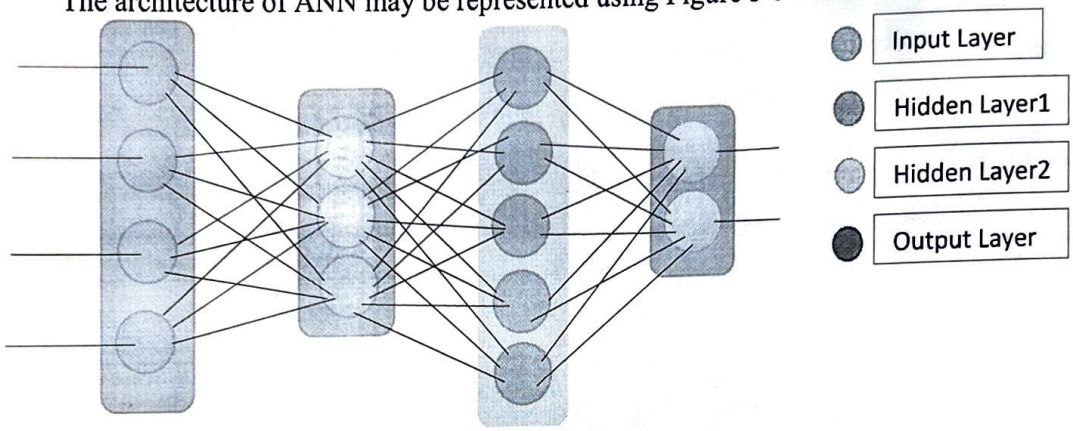

Figure 3: Architecture of ANN

b. Decision Tree (DT)

Decision Tree is supervised learning algorithm which is used for both classification and regression task. It is decision like model in which there are n number of possibilities and each possibility has further possibilities termed as conditions and based on all possible conditions of possibilities one action is performed. The following figure [figure 4] shows an example of how decision tree works:

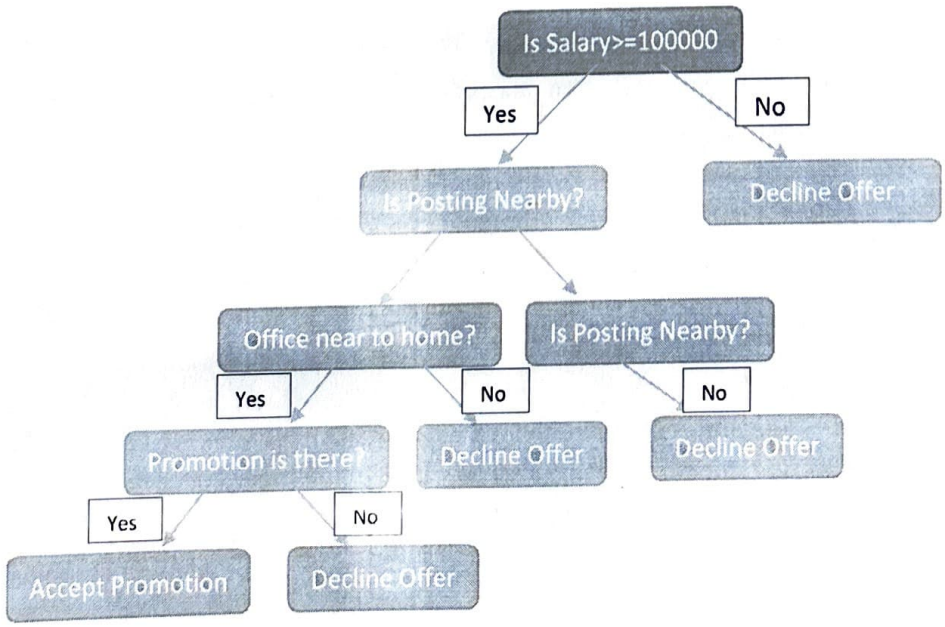

Figure 4: Working of Decision Tree

c. K-nearest Neighbour (KNN)

The K-nearest Neighbour (KNN) is a machine learning algorithm which is used for classification and regression. It is a simple machine learning algorithm which stores all the data and classify it in a new data point based on similarity. Once a new data point is created, the new data set may be analysed and placed on well suited category using KNN algorithm. Figure 5 below illustrates the concept of KNN Algorithm.

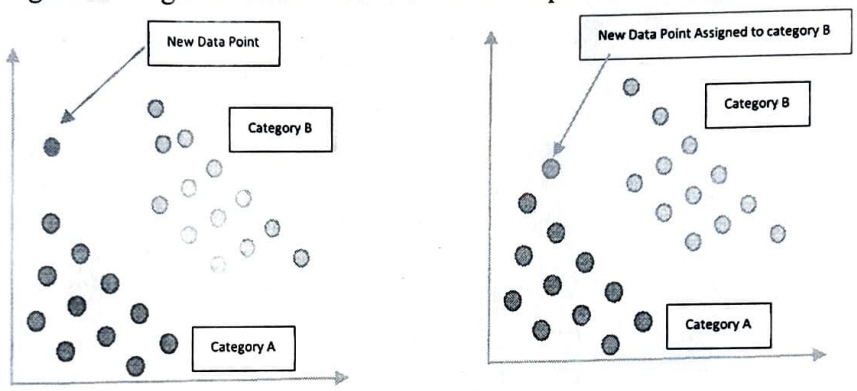

Figure 5 | KNN Algorithm

d. Support Vector Machine (SVM)

Support Vector Machine (SVM) is supervised ML algorithm which is used for complex dataset. It can be used for both classification as well as regression problems. In this algorithm margins are made using the points which are closest to hyperplane (support vetors). SVM can be understood using the following example:

Suppose there are two dataset of blue and yellow balls. Now it is to specify that whether the new ball will be of blue set or yellow set. The maximum margin, positive hyperplane, negative hyperplane and support vectors are taken into consideration for the final outcome.

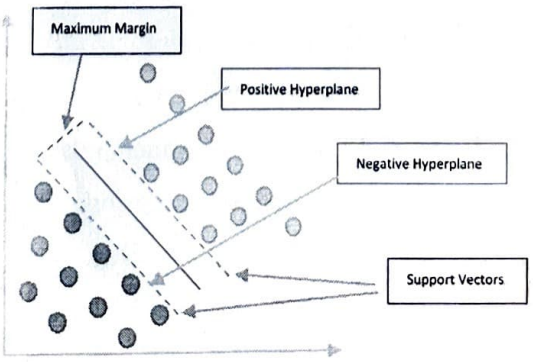

Figure 6: SVM Algorithm

e. Naïve Bayes

Naive Bayes is a supervised machine learning algorithm which is majorly used for classification purpose and more specifically text classification. It is a family of algorithm based on Bayes theorem. It is comparatively easy to implement and efficient for large number of features. It performs well even with limited number of training data. The working of Naïve Bayes Classifier may be represented through the figure 7 below.

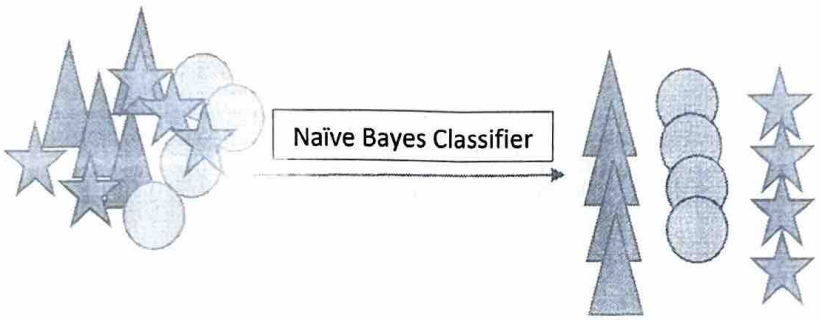

Figure 7: Naïve Bayes Classifier

5. Research Objectives

1. To study & analyse Machine Learning algorithms among diverse learners profile.
2. To identify & use emotion lexicons for emotion classification in online learning platform.
3. To develop an efficient emotion detection and classification technique.
4. To validate the proposed technique for emotional responses of students engaged in online course.

6. Research Methodology

Implementing emotion detection and responsive actions in the context of online courses can significantly enhance the learning experience for both educators and students. This initiative aims to facilitate smoother course completion by leveraging emotion detection techniques applied to textual data, particularly comments associated with course videos. The process will involve a specific sequence of steps to effectively identify and respond to various emotions expressed in the textual content. First of all the text is tokenized after which it is normalized and some special characters for example , . ‘ ‘ “ ” / ? [ ] () { } etc. are excluded from it. This process is called removing stop words. Further, POS (Parts of Speech) tagging is done which is a process to tag each word a parts of speech like noun, pronoun, verb, adverb, adjective etc.

After this each word is converted to its real form which is called lemmatization. Here, the raw data is converted in a format which could be further fed into machine learning algorithm. This process is known as Feature Extraction which may be done using TF-IDF method, word embedding, Ngram, Bag of words etc. This data is provided to the various machine learning algorithms to analyse. After applying various machine learning algorithm, each algorithm generates the results and the results are studied to get the desired result. Figure 8 as below shows the graphical representation of research methodology:

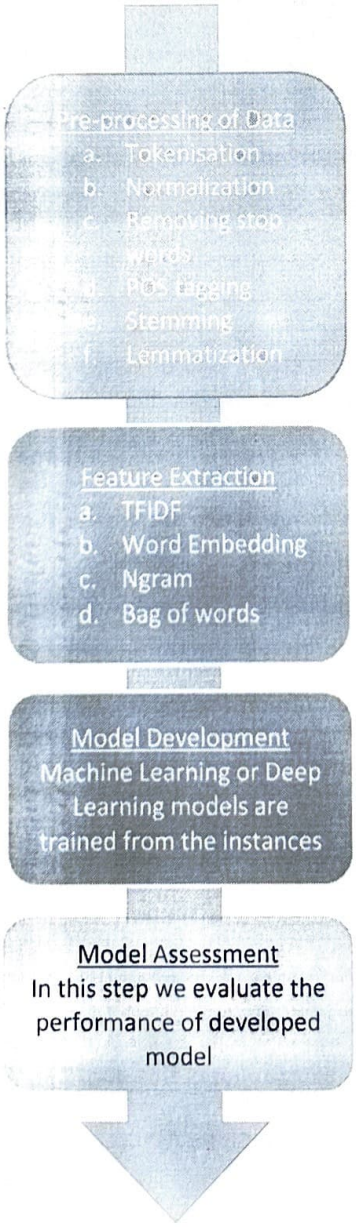

Figure 8: Research Methodology (Source: Self)

7. Expected Outcome of the Research

The outcome of this research will be to develop an emotion detection system to significantly improve the educational experience by providing timely emotional support specifically designed for students who are taking online courses. The insights gained from this research can help optimize emotional awareness in online learning environments.

References

[1] Abdi, A., Sedrakyan, G., Veldkamp, B. *et al.* Students feedback analysis model using deep learning-based method and linguistic knowledge for intelligent educational systems. *Soft Comput* 27, 14073–14094 (2023). https://doi.org/10.1007/s00500-023-07926-2

[2] Abdul Jalil, Nasir & Hwang, Ha & Mat Dawi, Norazryana. (2019). Machines Learning Trends, Perspectives and Prospects in Education Sector. ICEMT 2019: Proceedings of the 2019 3rd International Conference on Education and Multimedia Technology. 201-205. 10.1145/3345120.3345147.

[3] Machová K, Szabóova M, Paralič J and Mičko J (2023) Detection of emotion by text analysis using machine learning. *Front. Psychol.* 14:1190326. doi: 10.3389/fpsyg.2023.1190326

[4] Santosh Kumar Bharti, S Varadhaganapathy, Rajeev Kumar Gupta, Prashant Kumar Shukla, Mohamed Bouye, Simon Karanja Hingaa, Amena Mahmoud, "Text-Based Emotion Recognition Using Deep Learning Approach", *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 2645381, 8 pages, 2022.

[5] Du Y, Crespo RG, Martínez OS. Human emotion recognition for enhanced performance evaluation in e-learning. Progress in Artificial Intelligence. 2023 Jun;12(2):199-211.

[6] Demir E, Bilgin M. Sentiment Analysis from Turkish News Texts with BERT-Based Language Models and Machine Learning Algorithms. In2023 8th International Conference on Computer Science and Engineering (UBMK) 2023 Sep 13 (pp. 01-04). IEEE.

[7] Edalati, M., Imran, A.S., Kastrati, Z., Daudpota, S.M. (2022). The Potential of Machine Learning Algorithms for Sentiment Classification of Students' Feedback on MOOC. In: Arai, K. (eds) Intelligent Systems and Applications. IntelliSys 2021. Lecture Notes in Networks and Systems, vol 296. Springer, Cham. https://doi.org/10.1007/978-3-030-82199-9_2

[8] Estrada, María Lucía Barrón and Cabada, Ramón Zatarain and Bustillos, Raúl Oramas and Graff, Mario: Opinion mining and emotion recognition applied to learning environments, Expert Systems with Applications, (2020) doi:10.1016/j.eswa.2020.113265

[8] J. Wu, "Machine Learning in Education," 2020 International Conference on Modern Education and Information Management (ICMEIM), Dalian, China, 2020, pp. 56-63, doi: 10.1109/ICMEIM51375.2020.00020.

[9] Jan Philip Bernius, Stephan Krusche, Bernd Bruegge, Machine learning based feedback on textual student answers in large courses, Computers and Education: Artificial Intelligence, Volume 3, 2022, 100081, ISSN 2666-920X[9] Students feedback analysis model using deep learning-based method and linguistic knowledge for intelligent educational systems | Asad Abdi etc. | 2023

[10] Lwin HH, Oo S, Ye KZ, Lin KK, Aung WP, and Ko PP "Feedback analysis in outcome base education using machine learning," In: 2020 17th international conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON), 2020: IEEE, pp. 767–770

[12] M.Kaplan, Andreas and Haenlein, Michael: Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster, Business Horizons 441-450, (2020)

[13] Misuraca M, Scepi G, Spano M (2021) Using Opinion Mining as an educational analytic: an integrated strategy for the analysis of students' feedback. Stud Educ Eval 68:100979

[14] Mohamed, Yahia & Alkawsi, Gamal & Ali, Nor'Ashikin & Alhussian, Hitham & Bahbouh, Hussein. (2021). Predicting student's performance using machine learning methods: A systematic literature review. 357-362. 10.1109/ICCOINS49721.2021.9497185.

[15] Pujari, Yogendra Prasad & Prasad, Dumpa & Malleswari, Naga & Gupta, Neha. (2022). Implementation of Machine Learning Based Google Teachable Machine in Early Childhood Education. International Journal of Early Childhood Special Education. 14. 4132-4138.

[16] Sannyuya Liu, Shiqi Liu, Zhi Liu, Xian Peng, Zongkai Yang, Automated detection of emotional and cognitive engagement in MOOC discussions to predict learning achievement, Computers & Education, Volume 181, 2022, 104461, ISSN 0360-1315,

[17] Sultana, Faria. (2019). Grammar Translation Method (GTM) for young learners in learning L2: practice and certainty in Bangladesh.

[18] Tamrakar L, Shrivastava D and Ghosh DS "Student sentiment analysis using classification with feature extraction techniques," arXiv preprint arXiv:2102.05439, 2021

[19] Swathi G, Patnaik BM, Janani A, Karthik K, Divya J, Rao MJ. EMOTION RECOGNITION FROM TEXT USING MACHINE LEARNING.

[20] Van Nguyen K, Nguyen VD, Nguyen PX, Truong TT , and Nguyen, NL-T "UIT-VSFC: Vietnamese students' feedback corpus for, sentiment analysis," In 2018 10th International conference on, knowledge and systems engineering (KSE), 2018: IEEE, pp. 19–24.

[21] Xiaohui Tao, Aaron Shannon-Honson, Patrick Delaney, Christopher Dann, Haoran Xie, Yan Li, Shirley O'Neill, Towards an understanding of the engagement and emotional behaviour of MOOC students using sentiment and semantic features, Computers and Education: Artificial Intelligence, Volume 4, 2023, 100116, ISSN 2666-920X

[22] Bhattacharya S, Agarwala A, Roy S. Mood detection and prediction using conventional machine learning techniques on COVID19 data. Social Network Analysis and Mining. 2022 Dec;12(1):139.

[23] Alaa Alslaity & Rita Orji (2022): Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions, Behaviour & Information Technology, DOI: 10.1080/0144929X.2022.2156387

[24] Hadikhah Mozhdehi, M., Eftekhari Moghadam, A. Textual emotion detection utilizing a transfer learning approach. J Supercomput 79, 13075–13089 (2023). https://doi.org/10.1007/s11227-023-05168-5

[25] Vijayanand. G, Karthick. S, Hari. B, Jaikrishnan. V, 2020, Emotion Detection using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCICCT – 2020 (Volume 8 – Issue 08),

[26] Onyema EM, Almuzaini KK, Onu FU, Verma D, Gregory US, Puttaramaiah M, Afriyie RK. Prospects and Challenges of Using Machine Learning for Academic Forecasting. Comput Intell Neurosci. 2022 Jun 17;2022:5624475. doi: 10.1155/2022/5624475. PMID: 35909823; PMCID: PMC9337975.

[27] Zhai X and Lu M (2023) Editorial: Machine learning applications in educational studies. Front. Educ. 8:1225802. doi: 10.3389/feduc.2023.1225802

[28] Ahmad, S.F.; Rahmat, M.K.; Mubarik, M.S.; Alam, M.M.; Hyder, S.I. Artificial Intelligence and Its Role in Education. *Sustainability* 2021, *13*, 12902. https://doi.org/10.3390/su132212902

MANAV RACHNA
Vidyanatariksha

Dr. O.P. Bhalla Central Library
Manav Rachna International Institute of Research and Studies
(A deemed-to-be-university under section 3 of UGC Act 1956)

SIMILARITY VERIFICATION CERTIFICATE

Title: Emotion Detection using Machine Learning for Students Enrolled in Online Courses

Researcher: Anand kumar
Registration Number: 22/PHD/025

Supervisor: Dr. Rashmi Agrawal
Department: S.C.A.

This is to certify that the above mentioned Short Synopsis/Thesis was scanned for similarity index through Turnitin Software on 13/2/24 ....with submission ID 2392702083 and the similarity index was found to be 10 .....%.The complete report running into 28 ..Pages, 6034 ....Word count with File Size. 3406 ...KB is being forwarded to the Research Supervisor/DRC/DRB for review.

Date: 13-02-024

Rajesh Kumar 13/2/24
Dr. Rajesh Kumar
University Librarian
Dr. O.P. Bhalla Central Library
Manav Rachna International
Institute of Research & Studies
Faridabad

The complete report of the above Short Synopsis/Thesis has been reviewed and the similarity index is below the accepted norms as per the University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations 2018. The Short Synopsis/Thesis may be considered for further processing

Anand Kumar
Research Scholar

Research Supervisor

Emotion Detection using Machine Learning for Students Enrolled in Online Courses

by Anand Kumar

Submission date: 12-Feb-2024 01:59PM (UTC+0530)
Submission ID: 2292702083
File name: ng_Machine_Learning_for_Students_Enrolled_in_Online_Courses.docx (381.35K)
Word count: 6034
Character count: 34024

Emotion Detection using Machine Learning for Students Enrolled in Online Courses

SHORT SYNOPSIS

by

Anand Kumar
22/PHD/025

Under the supervision of
Supervisor
Dr. Rashmi Agrawal
Professor, School of Computer Application

SCHOOL OF COMPUTER APPLICATION
MANAV RACHNA INTERNATIONAL INSTITUTE OF RESEARCH & STUDIES
SECTOR 43, SURAJKUND – DELHI ROAD, FARIDABAD 121001

ABSTRACT

Soon after the COVID-19 outbreak, the entire education industry faced a whole new challenge for creating a secondary type of education that could be feasible, accessible, keeping in view the imposed restrictions. Online education in its various forms was introduced, and all of a sudden, the whole industry shifted to online education. The Universities/Colleges/Institutions were all of a sudden compelled to adopt online learning. Along with this, many online courses came into existence and mainly in the form of "Recorded Video" in which the student has to go through a self-paced course. MOOCs and various online courses were introduced and the same is being continued with a vast number of students enrolling in this. The study of the effectiveness of these courses is one of the most important tasks, as its effectiveness and completeness will shape its future. The study of Students' Emotion through text has to be analyzed in order to get their view on the online course they have enrolled. As per one of the research "Only 10% of enrolled students in MOOC's courses completes the course". Thus it becomes necessary to study the reason behind it. The comments by the students below each video may be analyzed using various machine learning algorithms in order to analyze students' emotion and act accordingly. This

will also facilitate the instructor to improve themselves according to the outcome supplied by students in terms of text comments. Using various machine learning techniques the comments of the students may be analyzed for the effectiveness and efficiency of respective online course. There are different machine learning methods for emotion detection, namely: Random Forest (RF), Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree (DT) etc.

All of these machine learning algorithms, as mentioned above, analyze emotions in text form, extract the most representative set of features, and then perform classification or clustering to model an optimal approach.

The research will utilize cutting-edge machine learning algorithms, including natural language processing (NLP), to capture and interpret emotional cues exhibited by students.

Key objectives include the design and implementation of a robust data collection framework, the development of feature-rich models for emotion classification, and the validation of the proposed system's efficacy across diverse online learning platforms.

Keywords: Machine Learning, Online Learning, Machine Learning in Education

13/2/24

Emotion Detection using Machine Learning for Students Enrolled in Online Courses

ORIGINALITY REPORT

Dr. Rajesh Kumar 13/2/24
University Librarian
O.P. Bhalla Central Library
Manav Rachna International
Institute of Research & Studies
Faridabad

10%	6%	9%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

Exclude small matches less than 14 words

PRIMARY SOURCES

1	int-jecse.net Internet Source	1%
2	Sannyuya Liu, Shiqi Liu, Zhi Liu, Xian Peng, Zongkai Yang. "Automated detection of emotional and cognitive engagement in MOOC discussions to predict learning achievement", Computers & Education, 2022 Publication	1%
3	www.researchgate.net Internet Source	1%
4	research.utwente.nl Internet Source	1%
5	mediatum.ub.tum.de Internet Source	1%
6	Jan Philip Bernius, Stephan Krusche, Bernd Bruegge. "Machine learning based feedback on textual student answers in large courses", Computers and Education: Artificial Intelligence, 2022 Publication	1%

7	link.springer.com Internet Source	1 %
8	Xiaohui Tao, Aaron Shannon-Honson, Patrick Delaney, Christopher Dann, Haoran Xie, Yan Li, Shirley O'Neill. "Towards an understanding of the engagement and emotional behaviour of MOOC students using sentiment and semantic features", Computers and Education: Artificial Intelligence, 2023 Publication	1 %
9	ris.utwente.nl Internet Source	1 %
10	Submitted to Manav Rachna International University Student Paper	<1 %
11	www.scilit.net Internet Source	<1 %
12	"Mining Intelligence and Knowledge Exploration", Springer Science and Business Media LLC, 2023 Publication	<1 %
13	"Proceedings of Third International Conference on Communication, Computing and Electronics Systems", Springer Science and Business Media LLC, 2022 Publication	<1 %

14	Bhagya M. Patil, Vishwanath Burkpalli. "A Perspective View of Cotton Leaf Image Classification Using Machine Learning Algorithms Using WEKA", Advances in Human-Computer Interaction, 2021 Publication	<1 %
15	teveso.cz Internet Source	<1 %
16	Morais, Jefferson, Yomara Pires, Claudomir Cardoso Jr., and Aldebaro Klautau. "A Framework for Evaluating Automatic Classification of Underlying Causes of Disturbances and Its Application to Short-Circuit Faults", IEEE Transactions on Power Delivery, 2010. Publication	<1 %
17	"Artificial Intelligence and Smart Environment", Springer Science and Business Media LLC, 2023 Publication	<1 %
18	Asad Abdi, Gayane Sedrakyan, Bernard Veldkamp, Jos van Hillegersberg, Stéphanie M. van den Berg. "Students feedback analysis model using deep learning-based method and linguistic knowledge for intelligent educational systems", Soft Computing, 2023 Publication	<1 %

19	Yahia Baashar, Gamal Alkawsi, Nor'ashikin Ali, Hitham Alhussian, Hussein T Bahbouh. "Predicting student's performance using machine learning methods: A systematic literature review", 2021 International Conference on Computer & Information Sciences (ICCOINS), 2021 Publication	<1 %
20	ijrpr.com Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 14 words
Exclude bibliography	On		